EPRS
European Parliamentary Research Service

At a glance
31/03/2014

Big data: opportunities and privacy concerns

As datasets grow larger and increasingly complex, tools for processing them have become indispensable. However, their growing use has serious implications for privacy. Hence a heated debate has arisen over "big data", with contrasting voices stressing either the gains or the risks involved and the difficulties in striking the right balance between the two.

Sizeable datasets: processing and use

Big data is a catch-all phrase describing data collections which merge data from multiple sources into a single one, so large that it cannot be processed using standard techniques. Specific tools are thus needed to extract information and identify previously unseen patterns in "noisy" and often unstructured data. The essential part of such a process is known as data mining and consists, for example, in grouping similar elements or people (cluster analysis) or capturing the co-occurrence of items (association analysis).

The growth of digital data has accelerated exponentially in recent years. A 2012 United Nations report stated that between 2007 and 2020 the world's stock of data was expected to increase 44 times. More digitised data were created between 2012 and 2013 than in the rest of human history. A new unit of "yottabyte", corresponding to one septillion (10^{24}) bytes, has been introduced to describe huge datasets.

The availability of such a massive amount of data presents numerous opportunities for public and private bodies alike. If efficiently processed, they may serve, and have already served, to improve many areas of human activity among which healthcare, weather forecasting, urban planning, and crime fighting are cited most often. As it is impossible to predict all future uses of big data, many argue that it should be stored for very long, if not indefinite, periods of time.

Privacy threats

In 2012, the press reported on the behavioural tracking practices of a retail company which used its customers' purchasing history to identify pregnant women. The company sent coupons for baby products to these customers, some of whom had not yet revealed their pregnancy to their families. This widely discussed case is one of many examples of how advances in data analysis may result in what is likely perceived by the individuals concerned as intrusions in their privacy. It is suggested that privacy threats brought about by big data may be divided into three broad categories: surveillance, disclosure and discrimination.

Surveillance

The feeling of being constantly monitored can seriously affect people's behaviour, discouraging them from adopting new technologies and searching for or reading certain materials over the internet. Moreover, surveillance may have a "chilling effect" on free speech, as it leads to individuals suspecting that a trace of everything said is kept somewhere and may be used against them in the future.

Disclosure

Disclosure may result from a failure to protect a data collection from a security breach (e.g. through hacking) or from the initial collector revealing information to a third party (secondary usage). Some companies allow the transfer of their clients' data to commercial partners, as illustrated by banking practices recently reported by the Dutch and Belgian press. Moreover, whereas data are mainly collected and stored by private entities, government agencies reportedly enjoy easy access to them, even in the absence of a suspicion of crime, let alone a court order, and without the individuals concerned being informed. This suspected easy access became front-page news with Edward Snowden's revelations concerning a mass surveillance programme operated by the US National Security Agency (NSA).

Author: Piotr Bąkowski, Members' Research Service
European Parliamentary Research Service
140771REV1

Profiling and discrimination

People may be treated differently on the basis of information collected about them on account of a specific characteristic (which they might not be willing to reveal), such as religious convictions or sexual orientation. As a result, services could be refused to certain people or companies offer highly personalised pricing, which may amount to discrimination.

EU policies and laws

An opportunity for Europe

Most existing legislative and non-legislative instruments relevant to data in general and big data in particular are part of the EU's open data policy. This policy focuses on the availability of information produced, collected and commissioned by the public sector in the EU. The legal framework in this field is provided by the 2003 Directive on the re-use of public sector information, substantially amended in 2013. The European Union Open Data Portal has been developed as the single point of access to data from EU institutions and other bodies. Moreover, relevant actions at national level are promoted and funded by the EU e.g. through the Seventh Framework Programme for research.

When addressing big data, the EU institutions are however aiming well beyond open data. As evidenced by a recent speech by the Commission's Vice-President Neelie Kroes, EU leaders have recognised big data's potential for sparking technological innovation, creating new jobs and building up the knowledge-based economy. The Commission is developing a strategic initiative based on the concept of the data value chain, which refers to the life-cycle of data (generation, processing and use in new innovative products and services). The October 2013 European Council Conclusions confirmed a comprehensive approach to big data, linking it to developments in cloud computing and supporting the creation of a single market for big data.

Data protection issues

Taking a strict position on privacy could arguably lead to restraining potential developments in science, which derives valuable insights from big data. Therefore calls have been made for a different approach, that of data empowerment – vesting individuals with the power to use their data in ways they wish – which could still be reconciled with the general public's interest in processing the data concerned.

However, existing laws need to be respected and the EU, compared to other parts of the world, has particularly strong data protection rules. The right to protection of personal data is a fundamental right, enshrined in Article 8 of the legally binding Charter of Fundamental Rights of the EU. The 1995 Data Protection Directive – which, in the absence of a compromise on the data protection reform package, remains the key EU legal instrument in this field – provides for a series of specific privacy principles including:

- **Purpose specification and limitation**: data are to be collected for specified, explicit and legitimate purposes and cannot be further processed in a way incompatible with these. However, in general, further processing for historical, statistical or scientific purposes should not be considered incompatible with the initial purpose, provided that Member States ensure appropriate safeguards.
- **Transparency** and **consent**: individuals are to be informed about the processing of their data, which requires explicit consent that can be withdrawn at any time.
- **Data minimisation**: data must be adequate, relevant and not excessive in relation to the purposes for which they are processed;
- **Limited retention**: keeping data in a form which permits identification of a person only as long as it is necessary for the purposes for which the data were collected or for which they are further processed.

The Directive equally protects individuals whose personal data are clearly linked to their name in the dataset concerned and those described in a way which enables identification through further research. The latter situation has become more common, with big datasets combining information from a variety of sources.

The robustness of the EU framework aside, it is important to note that the right to data protection has its limits. According to the Court of Justice of the EU the right to data protection is not "an absolute right, but must be considered in relation to its function in society", a position which underscores the need for a balanced approach when addressing opportunities and risks of big data analysis.